بسم الله الرحمن الرحيم

محركات البحث

(Search Engine)

تأليف:

م.تقني: عبدالرحمن محجوب حمد

Introduction

الحمد لله بنعمته تتم الصالحات و بشكره تزداد الخيرات والبركات ها انا اضع هذا الكتاب خالص لوجه الله تعالى يحتوى علـــى معلومـــات مهمة بالنسبة لمستخدمي الانترنت و متطوري المواقع بلغة HTML يحتوي الكتاب على تعريف بمحركات البحث وانواعه و آلياته والفرق بينه وبين الادلة......!

تعريف بالمؤلف:

الاسم:عبدالرحمن محجوب حمد.

العنوان: السودان.

تاريخ الميلاد:١٩٨٦م.

المؤهلات:دبلوم تقني ثلاث سنوات.- في قسم تقنية علوم الحاسوب.

المهارات:احب تطوير المواقع عن طريق لغـــاتHTML و PHP و asp وتصميم الانظمة بلغة visual basic و oracle.

البريد الالكتروني:a_mahagob@yahoo.com

التلفون:٠٠٢٤٩٩١٩٥٧٤٥٧٣

والله من وراء القصد،،،

محركات البحث

Search Engine

محركات البحثSearch Engine هي أدوات شائعة الاستخدام إلى الوصول إلى صفحات النسيج الشبكي العالمي world wide web لكنها تعود بمئات النتائج والوصلات التي تكون في الغالب غير ذات صلة بموضوع البحث الرئيسي مما يجعل الباحث في حيرة ويأس من وجود ما يبحث عنه، فإن استخدام أدوات البحث المتوفرة على الإنترنت بدون استراتيجية بحث محددة تشبه إلى حد كبير من يسير في مكتبة ضخمة بشكل عشوائي يحاول العثور على كتاب معين.

وسوف يناقش هذا البحث النقاط التالية:

١- الفرق بين محركات البحث والأدلة .
٢- مكونات محركات البحث على شبكة الإنترنت وكيفية عملها.
٣- نسبة المواقع التي تغطيها محركات البحث من مجموع مواقع الشبكة.
٤- فهرسة محركات البحث لمواقع الشبكة.
٥- الوقت اللازم لظهور صفحات الويب الجديدة في محركات البحث.\
٦- آليات البحث في محركات البحث.
٧- أنواع محركات البحث.

أولا : الفرق بين محركات البحث والأدلة :

يسخر مستخدمو الإنترنت أدوات البحث المختلفة على الوب للوصول إلى المعلومات المفيدة الموزعة على ملايين الأجهزة الخادمة على الإنترنت، وهذه الأدوات هي محركات البحث search engine والأدلة directories التي تساعد في تسريع عملية البحث بدرجة كبيرة، وفي هذا السياق لابد تقديم تعريف مبسط لهاتين الأداتين قبل أن دخل في تفاصيل محركات البحث.

محركات البحث Search Engines:

وهي يمكن تعريفها بأبسط صورة هي عبارة عن قواعد بيانات ضخمة بعناوين ومواقع، ومع وصف مصغر لصفحات الإنترنت المختلفة، والتي بواسطتها أي محركات البحث يمكن البحث ع موضوع معين في حقل من الحقول المختلفة في الشبكة بشكل دائم بغرض إيجاد دليل معين لمثل هذه الصفحات، ولأنها تعمل بشكل آلي وتقوم بفرز وفهرسة كم هائل من الصفحات ، فسيلاحظ كل شخص يستخدم هذه المحركات على أنها تحتوي على كثير من المعلومات غير المتوفرة في الأدلة directories وقد يمثل ذلك جانب إيجابي يجعل من محركات البحث أداة فعالة أكثر من الأدلة

الأدلة Directories :

الأدلة لا تعمل بشكل آلي بل تتم إدارتها من قبل أشخاص متخصصين، وما يحدث هو أن العديد من المواقع يتم تسليمها إلى دليل ما، ومن ثم يتم فرزها وتبويبها تحت تصنيف معين، ولأن هذه الآله يتم إدارتها بشكل بشري فإنها قادرة دوما على توفير معلومات أكثر دقة وموقع yahoo المشهور ما هو إلا مثال واحد لهذه الأدلة

وتقدم الأدلة للمستخدم طريقة سريعة للبدء بعمليات البحث عن المعلومات بواسطة تفحص المواضيع المصنفة التي يعرضها ، إذ يندرج تحت كل موضوع لائحة من المواضيع الفرعية فيمكن للمستخدم أن يتفحصها تباعا إلى أن يصل إلى المعلومات المطلوبة، وفي حال عدم وجود المعلومات تحت الموضوع الذي اختاره المستخدم، يتراجع ويختار موضوعا رئيسيا آخر ليقوم بالبحث في تفرعاته من جديد، وهكذا...

وفي هذا السياق لابد من توضيح من أن هناك العديد من أدلة البحث تعمل أيضا كمحركات منها : yahoo – altavista – excite

ثانيا : مكونات محركات البحث وكيفية عملها :

يوجد على شبكة الإنترنت كميات ضخمة من المعلومات تتوزع على شكل صفحات وتنمو بشكل سريع (في عام ١٩٩٦ بلغ معدل النمو ٤٠٠ %)، و في أوائل ١٩٩٨ كان هناك ما يتعدد بـ ٣٠٠ مليون صفحة متوفرة على النسيج، وفي أواخر اعم ٢٠٠٠ قدر عدد الصفحات بأكثر من مليار صفحة، وهو الآن يقدر بأن يكون حوالي مليار ونصف في بداية عام ٢٠٠٢، ويقدر عدد الصفحات التي تضاف أو تحدث يوميا بمليون ومائتين صفحة، وهذا يدل على معدل نمو الإنترنت الذي يتضاعف بكل سنوي.

هناك أكثر من ٢٥٠٠ أداة بحث متوفرة ويقدر تغطية تلك الأدوات للإنترنت بـ ٥ % للأدوات الصغيرة و ١٥ % لتلك العملاقة، ومن هنا نعلم أنه مهما كان سمعة أداة البحث فإنها لا يمكن أن تغطي كافة المعلومات المتوفرة على الشبكة، وتظهر محركات البحث نتائج مختلفة، لأن كل منها يبحث في قاعدة بياناته الخاصة به، التي تتضمن جزء من مواقع الوب ، ويتألف محرك البحث من ثلاثة أجزاء رئيسية هي:

١. برنامج العنكبوت Spider Program
٢. برنامج المفهرس Indexer Program
٣. برنامج محرك البحث

١. برنامج العنكبوت

تستخدم محركات البحث برنامج العنكبوت لإيجاد صفحات جديدة على الوب إضافتها، ويسمى هذا البرنامج أيضا الزائر لأنه يبحر في الإنترنت بهدوء، لزيارة صفحات الويب والإطلاع على محتويتـــها، ويأخذ هذا البرنامج مؤشرات المواقع من عنوان الـــصفحة Page Title والكلمـــات المفتاحيـــة Keywords التي تحتويها ، إضافة إلى محتويات محددات الميتا Meta Tags فيها.

ولا تقتصر زيارة برنامج العنكبوت على صفحات الموقع، بل يتابع البرنامج تعقب الروابط (links) الموجودة فيها لزيارة صفحات أخرى. أما الغاية من هذه الزيارات فهي وضع النصوص المتاحة علـــى تلك المواقع على فهارس محرك البحث ليتمكن المحرك من العودة إليها فيما بعد، و لم تغب فكرة تغـــيير المحتوى في الموقع عن البال مصممي محرك البحث، إذ يقوم محرك البحث بزيارات دوريـــة للمواقـــع الموجودة في الفهارس للتأكد من التعديلات التي تصيب المواقع المفهرسة.

٢. برنامج المفهرس:

يمثل برنامج المفهرس index program ،ويطلق عليه أحيانا الكتالوج catalogue ، قاعـــدة بيانات ضخمة توصيف صفحات الويب ويركز التوصيف على المعلومات التي حصلت عليهـــا مـــن برنامج العنكبوت spider كما تعتمد على بعض المعايير مثل الكلمات الأكثر تكرارا من غيرهـــا، وتختلف محركات البحث عن بعضها في هذه المعايير، إضافة إلى اختلافها في خوارزميـــات البحـــث searching algorithms

٣. برنامج محرك البحث:

يبدأ دور برنامج محرك البحث search engine program عند كتابة كلمة مفتاحية في مربع البحث search box إذ يأخذ هذا البرنامج الكلمة المفتاحية ويبحث عن صفحات الويـــب الـــتي تحقق الاستعلام، الذي كونه برنامج المفهرس في قاعدة بيانات الفهرس، ثم يعرض نتيجـــة البحـــث المتمثلة بصفحات الويب التي طلبها المستخدم في نافذة المستعرض browser window ويقـــوم أيضا بعملية الترتيب لهذه الصفحات

وتختلف محركات البحث عن بعضها في أسلوب العمل، فمثلا: تحتفظ قاعدة بيانـــات altavista بكل تفاصيل صفحة الويب المخزنة عليها، أما محركات البحث الأخرى فقد تحتفظ بالعناوين الرئيسية للصفحة فقط ، مما يؤدي إلى اختلاف شكل ودقة نتائج البحث الظاهر للمستخدم.

ثالثا : نسبة المواقع التي تغطيها محركات البحث من مجموع مواقع الشبكة :

أكدت دراسة أجراها معهد NEC للبحوث الواقع في مدينة برينيتون في ولاية نيوجرسي الأمريكية عام ١٩٩٩، على ١١ محرك بحث مستخدمة في الإنترنت، إن هذه المحركات لم تتمكن من التأقلم مع النمو الكبير لمواقع الويب، وبينت الدراسة أن محركات البحث لا تغطي سوى قـــسم ضـــئيل مـــن

صفحات الويب، وان إدراج الصفحات الجديدة بينها يستغرق وقتا طويلا. ووجـــدت الدراســـة أن محرك البحث northern light هو أكثر المحركات شمولا لكنه على الرغم من ذلـــك لا يغطـــي سوى ٦/١ من صفحات الإنترنت التي يمكن أن تصل الإنترنت إليها، ويذكر أن هذه النسبة قبل ١٨ شهر كانت تصل إلى ٣/١ من صفحات الإنترنت. ويقول ستيف

لورانس وسي لي جاليز، المتخصصان في علوم الكمبيوتر في معهد NEC والقائمين بهذه الدراسة، أنه من المحتمل أن يكون تأخر إدراج الصفحات الجديدة ضمن محركات البحث أمـــرا مقـــصودا، إذا تمتلئ قواعد البيانات التابعة لمحركات البحث هذه وأصبحت عملية إضافة موقع جديد إليهـــا عمليـــة مكلفة، خاصة إن هذه الإضافة لا تزيد في معظم الأحوال من الأرباح الإعلانية.

أسم محرك البحث	نسبة التغطية البحث	دعم اللغة العربية	عنوان محرك
SNAP	%١٥,٥	لا	WWW.SNAP.COM
ALTAVISTA	%١٥,٥	نعم	WWW.ALTAVISTA.COM
HOT BOT	%١١,٣	لا	WWW.HOTBOT.COM
MICROSOFT	%٨,٥	لا	WWW.MSN.COM
INFOSEEK	%٨,٥	نعم	WWW.INFOSEEK.COM
GOOGLE	%٧,٨	نعم	WWW.GOOGLE.COM
YAHOO	%٧,٤	لا	WWW.YAHOO.COM
LYCOS	%٢,٥	لا	WWW.LYCOS.COM
EUROSEEK	%٢,٢	نعم	WWW.EUROSEEK.COM
EXICIT	%٥,٦	نعم	WWW.EXCIT.COM
ALL THE WEB	١٥,٥	نعم	WWW.ALLTHEWEB.COM

رابعا : فهرسة محركات البحث للمواقع :

ومن خلال العرض السابق لمكونات محركات البحث، تبين أن برنامج العنكبوت أو عنكبوت الشبكة " web spider وبرنامج المفهرس cataloger هما المسئولان عن إيجاد صفحات الويب الجديدة وفهرستها وإدخالها داخل قاعدة بيانات المحرك ولكن كيف ؟

تستخدم واصفات <META> الخاصة بكلاً من وصف الموقع والكلمات المفتاحية في صفحة HTML ، حتى تضمن إدراج المواقع بشكل سليم في محركات البحث، ووضع وصف واضح له فيها، ولا يؤدي استخدام واصفة <META> بعكس كثير من واصفات HTML إلى ظهور أي عنصر مرئي على صفحتك، لكن عمل هذه الواصفة يتم في الكواليس، حيث تقدم وصفاً ن المعلومات التي تتضمنها الصفحة إلى المتصفحات وعناكب الشبكة ومحررات HTML .

إذا الوظيفة الأساسية هي وصف الصفحة وكلمات المفتاح فيها وتقديمها لمحركات البحث. عندما يزور عنكبوت الشبكة صفحة لا تحتوي على وصف لها باستخدام واصفة <META> فإنه يسجل الأسطر القليلة الأولى من النص الموجود في بداية الصفحة ويعتبر ذلك وصفا لها ، ولا يعطي هذا الأسلوب غالبا وصفاً دقيقاً للصفحة. وأحد الأسباب الرئيسية لذلك "وجود ملف صورة في بدايتها"، إذ لا يتعرف عنكبوت الشبكة إلا على النصوص، ولا يستطيع قراءة الملفات الرسومية، التي تتضمن نصوص على هيئة صور، وهذا يوضح استخدام واصفات <META> وخاصتاً عندما تحتوي الصفحة على كثير من الصور.

خامسا : الوقت اللازم لظهور صفحات الويب الجديدة في محركات البحث:

من خلال بحث أجرته مجلة الطبيعة الأمريكية في عدد يوليو من عام ١٩٩٩ والتي أوضحت أن المتوسط العمري لظهور الموقع في محركات البحث هو ٥٧ يوم تقريبا وحاز محرك البحث ALTAVISTA على المركز الأول فيما بينها والإحصائيات كانت على النحو التالي:

Altavista	٣٣ يوم
Excite	٤٧ يوم
Northen Light	٤٨ يوم
Hot Bot	٥١ يوم
MSN	٥٧ يوم
Infoseek	٦٠ يوم
Yahoo	٧٦ يوم
Snap	٩١ يوم

Lycos	١٤٧ يوم
Uroseek	غير معروف
Google	غير معروف

كما أوضحت الدراسة أن الوقت اللازم لتحديث قاعدة بيانات كل محرك بحث هي نفس المعدلات السابقة، حيث لا بد أن يتم تحديث قواعد بيانات محركات البحث لملاحقة التغيرات أو التعديلات على المواقع الموجودة داخل قاعدة البيانات وكذلك الحصول على مزيد من المواقع الجديدة على الويب.

سادساً : آليات البحث في محركات البحث:

ذكرنا سابقا أن محركات البحث تستخدم في بحثها عن مواقع الويب ما يدعى الكلمات المفتاحية Keywords التي يمكن ان تكون عبارة، وتستخدم آليات البحث عادة بعض المعاملات Operators مع هذه الكلمات المفتاحية، لتوفير خيارات إضافية لعملية البحث، وهناك طريقتين للبحث في محركات البحث هما:

البحث البسيط SIMPLE SEARCH

وهذا النوع من أنواع البحث يقوم به معظم الناس وخاصتا المبتدئين مع استخدام الإنترنت وهم من يجهلون تقنيات البحث المتقدم. وهو يكون عن طريق وضع كلمة بدون أي علامات أو شارات ثم البحث عنها ولكننا سنجد النتيجة العديد من النتائج العلائقية والروابط المتينة.

البحث المتقدم ADVANCED SEARCH

تزيد خصائص البحث المتقدمة التي يوفرها الكثير من محركات البحث إمكان العثور على المعلومات، كما إنها طريقة فعالة للبحث عن معلومات محددة إذ تتيح للمستخدمين إمكان البحث عن عدة كلمات مفتاحية معاً.

سابعا : أنواع محركات البحث :

يمكن القول أنه قد ولى الزمان الذي كان يمكن فيه الاكتفاء بموقع بحث واحد ليلبي مختلف احتياجاتك في البحث بالشبكة، فمع التنامي المطرد للإنترنت جرى نوع من التخصص في وظائف البحث، بحيث أصبح كل موقع لمحركات البحث يحتوي على مجموعة من الخدمات التي تحتاجها للتنقيب عن المعلومات المطلوبة بسرعة ويسر، وسيتم عرض تلك الأنواع وهي كالآتي :

١. **محرك بحث يبحث عن الصور** :يمكن القول بأن موقع www.ditto.com وهو محرك بحث يرتكز على الصور، من إدخال كلمة، أو قطع للبحث عنه، وبدلا من إعادة نتائج البحث

بشكل نصي يعيد الموقع النتائج بشكل عدد من الصور المصغرة Thumbnails ، وعند النقر على أي صورة مصغرة تعرض صفحة الويب التي تتضمنها، ويضم الفهرس في هذا الموقع أكثر من مليوني صورة، وهذا طبقا لإحصائية تم عملها في عام ١٩٩٩، كما أن هناك العديد من محركات بحث الصور منها www.snap.com

٢. محرك بحث الملفات:

قامت شركة c2v بتصميم برنامج جديد أطلق عليه أسمtextomattom وهو من محركات البحث متعددة اللغات وليس من بينها العربية، وهو يعمل على إيجاد الملفات الهامة عن طريق محتوياتها أيا كانت نوعها أو مواقعها على الويب، كما يعمل على تنفيذ عمليات البحث اعتمادا على إدخال كلمات مفاتيح المفردات المتخصصة، وبالتالي يستطيع البرنامج من خلال المحتوى دون اللجوء إلى عناوين الملفات ودون الحاجة إلى فتحها أياُ كان نوع الإصدارة، وكذلك يسمح باستخدام كلمات البحث المتقدم and , or , not .

٣. محركات بحث متخصصة:

كما أن هناك محركات بحث عامة ، فهناك محركات بحث متخصصة في موضوعات معينة مثل :

- محركات بحث طبية: www.nlm.nih.gov و www.healthfinder.org
- محركات بحث في مناطق أو دول معينة : مثل افريقيا www.rubani.com . و www.ananzi.com

الخاتمة

الحمد الله الذي بنعمته تتم الصالحات الصلاة والسلام على المصطفى نبيه.....الحمد لله ايقظ روح الامل والتفاؤل من جديد في كل دقة تنعكس على مسار العمـــل وايجـــاد المختصر المفيد.هذا الكتاب بذلت فيه ما وفقني الله تعالى إليه ، وســـلكت في كتابتـــه أسلوبا سهلا مع تنسيقا جيدالسهولة تصفحه ، واستخدمت ألفاظا سهلة المنال.

وهذا الكتاب أؤكد أنه لمن لا يعرف شيئا عن محركات البحث ،سوف يجده هنا اضـــافة الي الادلة.ولا تنسوني إخواني من دعوة صالحة عن ظهر غيب عسى أن يغفر الله لي بها تقصيري ،

والله ولي التوفيق ،،،،

م/تقني:عبدالرحمن محجوب حمد
a_mahagob@yahoo.com